

مقایسه ساختار اصطلاحنامه‌های پایگاه‌های اطلاعاتی Pubmed و Embase با استاندارد اصطلاحنامه نویسی سازمان ملی استاندارد‌های اطلاعاتی آمریکا و بررسی شیوه‌های نمایه سازی دو پایگاه

صدیقه خانی^۱ / محمد رضا علی بیگ^۲ / حمید حقانی^۳

چکیده

مقدمه: طبق آمار مرگ و میر سال ۱۳۸۲، بیماری‌های قلبی و عروقی، سرطان‌ها، حول تولد و دستگاه تنفس بیشترین مرگ و میر را در ایران داشته‌اند. برای کاهش میزان مرگ و میر پزشکان ایرانی نیازمند کسب آگاهی از آخرین روش‌های درمانی هستند. دو مورد از مهمترین پایگاه‌های علوم پزشکی و پرکاربردترین آنها Pubmed و Embase می‌باشند. شناسایی روش‌های نمایه سازی و مقایسه اصطلاحنامه‌های دو پایگاه اطلاعاتی Pubmed و Embase، به جستجوی موفق تر اطلاعات کمک می‌کند و به دنبال آن می‌توان الگویی روزآمدتر برای شیوه‌های نمایه سازی و اصطلاحنامه پایگاه ملی اطلاع رسانی پزشکی بدست آورد. لذا این مطالعه با هدف مقایسه شیوه‌های نمایه سازی Pubmed و Embase انجام گردید.

روش بررسی: این پژوهش از نوع کاربردی و روش پژوهش پیمایشی تحلیلی است. در ۳ ماهه چهارم سال ۲۰۰۶، برای ارزیابی اصطلاحنامه‌ها تمامی توصیفگرهای مربوط به ۴ زمینه موضوعی از دو اصطلاحنامه استخراج گردید و برای شیوه‌های نمایه سازی با طراحی راهبرد جستجویی مقالات جمع آوری شد و ۶۳۸۱ توصیفگر و ۳۳۵۸ مقاله ارزیابی گردید. برای جمع آوری داده‌ها از لیست کنترل و پرسشنامه استفاده شد. نه فاکتور از دو اصطلاحنامه (شامل توصیفگر عبارتی در برابر توصیفگر تک کلمه، تعداد واژه‌ها در توصیفگر عبارتی، توصیفگر در قالب صفت و موصوف در برابر توصیفگر با ترکیب حرف اضافه و یا حرف ربط، توصیفگر غیر مقلوب در برابر توصیفگر مقلوب، استفاده از انواع علامت‌های نشانه گذاری در ساختار توصیفگر، استفاده از اختصار و آغازه به عنوان توصیفگر، حروف تعریف در توصیفگر، توصیفگر با توضیحگر، ارائه یا داشتن دامنه) با استاندارد مقایسه گردید. تحلیل داده‌ها با استفاده از نرم افزار SPSS و آزمون‌های آماری t و z انجام شد.

یافته‌ها: اصطلاحنامه Emtree در ۶ فاکتور و اصطلاحنامه MeSH در ۴ فاکتور انطباق بیشتری با استاندارد داشته‌اند. بیشتر مقالات در پایگاه اطلاعاتی Pubmed با ۲۱-۳۰ واژه و در پایگاه اطلاعاتی Embase با ۳۱-۴۰ واژه نمایه سازی می‌شوند.

نتیجه گیری: ساختار Emtree برای الگوبرداری مناسب تر از MeSH است. شیوه نمایه سازی Embase تخصیصی، استخراجی، جامع و اخص نگر است. در Pubmed نمایه سازی به صورت تخصیصی و جامع صورت می‌گیرد.

کلید واژه‌ها: ارزیابی اصطلاحنامه‌ها، روش‌های نمایه سازی، Embase, Pubmed, Emtree, MeSH

◇ وصول مقاله: ۸۶/۷/۱۷، اصلاح نهایی: ۸۶/۲/۱۸، پذیرش مقاله: ۸۶/۲/۲۵

۱- کارشناس ارشد کتابداری و اطلاع رسانی، دانشکده مدیریت، دانشگاه علوم پزشکی ایران

۲- مربی گروه کتابداری، دانشکده مدیریت، دانشگاه علوم پزشکی ایران، نویسنده مسئول (Email: simakhib@yahoo.com)

۳- مربی گروه آمار، دانشکده مدیریت، دانشگاه علوم پزشکی ایران

مقدمه

بر طبق آمار سیمای مرگ و میر در ۲۳ استان کشور سال ۱۳۸۲، به ترتیب بیماری‌های قلبی عروقی، سرطان‌ها، بیماری‌های حول تولد و بیماری‌های دستگاه تنفسی بالاترین میزان مرگ و میر را در ایران داشته‌اند. [۱] جامعه پزشکی ایران که با بیماری‌های فوق سروکار دارند برای کاهش میزان مرگ و میر ناشی از آنها نیازمند کسب آگاهی از آخرین روش‌های درمانی هستند. به طور مداوم در علم پزشکی برای درمان بیماری‌ها روش‌های گوناگونی معرفی می‌شود که راه‌حل‌های پیشرفته و موثرتری را برای بیماری‌ها ارائه می‌دهند. نتایج تحقیقات فوق بیشتر در قالب مقالات مجلات به جامعه پزشکی معرفی می‌شوند. پایگاه‌های اطلاعاتی علوم پزشکی مقالات را جمع‌آوری کرده آنها را نمایه‌سازی و چکیده‌نویسی می‌کنند تا کاربران بتوانند به راحتی اطلاعات مورد نیاز خود را بازیابی کنند.

دو مورد از شناخته شده ترین پایگاه‌های اطلاعاتی علوم پزشکی که صرفاً در این زمینه فعالند، Pubmed و Embase هستند. Pubmed، نخستین پایگاه اطلاعاتی است که اطلاعات تولید شده در علوم پزشکی را جمع‌آوری و سازماندهی می‌کند و به صورت رایگان و از طریق اینترنت در دسترس عموم قرار می‌دهد. [۲] Embase، پایگاه اطلاعاتی Science Direct است و بر خلاف Pubmed هزینه اشتراک بسیار بالایی دارد. [۳] اسدی اربابی رفتارهای اطلاع‌یابی محققان مراکز تحقیقاتی علوم پزشکی شهر تهران را بررسی کرده است. نتایج پژوهش نشان می‌دهد، بیشترین میزان مراجعه افراد به پایگاه‌های اطلاعاتی، مربوط به Pubmed / Medline است و Excerpta Medica / Embase در مرتبه سوم قرار دارد. [۴] نمایه‌سازی (Indexing) در پایگاه اطلاعاتی، مسئله اصلی است. [۵] شیوه‌های مختلف نمایه‌سازی و اصطلاحنامه‌ای که در نظم‌دهی به مطالب پایگاه اطلاعاتی به کار می‌رود، از عوامل تعیین‌کننده کیفیت نمایه‌سازی هستند. پزشکان برای بازیابی اطلاعات مورد نیاز خود، نیازمند جستجو در این پایگاه‌های اطلاعاتی هستند و آگاهی از شیوه‌های نمایه‌سازی می‌تواند به طراحی راهبرد جستجوی موفق‌تر و بازیابی راحت‌تر اطلاعات کمک کند. شناسایی شیوه‌های نمایه‌سازی به کار رفته در دو مورد از بزرگترین

پایگاه‌های اطلاعاتی علوم پزشکی و بررسی اصطلاحنامه‌های آنها، می‌تواند به عنوان الگویی برای نمایه‌سازی داده‌های پایگاه ملی اطلاع‌رسانی پزشکی و تدوین یک اصطلاحنامه پزشکی فارسی روزآمدتر به کار برده شود. لذا این مطالعه با هدف مقایسه شیوه‌های نمایه‌سازی Pubmed و Embase انجام گردید.

روش پژوهش

این پژوهش از نوع کاربردی و روش پژوهش پیمایشی تحلیلی است. جامعه پژوهش تمامی توصیفگرهای (Descriptors) اصطلاحنامه‌های MeSH و Emtree و نیز مقالات مجلات درباره بیماری‌های پرم‌رگ و میر است که در هر دو پایگاه اطلاعاتی Pubmed و Embase نمایه‌شده‌اند.

۶۳۸۱ توصیفگر از شاخه‌های موضوعی نامبرده و در ساختار سلسله‌مراتبی دو اصطلاحنامه مورد بررسی قرار گرفت. ۲۴۶۸ توصیفگر در ساختار سلسله‌مراتبی اصطلاحنامه MeSH و ۳۹۱۳ توصیفگر در ساختار سلسله‌مراتبی اصطلاحنامه Emtree ارزیابی گردید. برای بررسی شیوه‌های نمایه‌سازی تعداد کل ۳۳۵۸ مقاله جمع‌آوری شد که ۱۶۷۹ مقاله از پایگاه اطلاعاتی Pubmed با ۱۶۷۹ مقاله از پایگاه اطلاعاتی Embase مقایسه شد.

نمونه‌گیری در سه ماهه چهارم سال ۲۰۰۶ انجام گرفت. برای مقایسه اصطلاحنامه‌ها با استاندارد از هر دو اصطلاحنامه تمامی توصیفگرهایی که در شاخه‌های موضوعی نامبرده بودند استخراج شد. برای اطمینان از جمع‌آوری تمامی توصیفگرها، توصیفگرهای ذخیره شده دوباره با اصطلاحنامه‌ها مقابله گردید.

ساختار اصطلاحنامه از اجزای متفاوتی تشکیل شده است که در این پژوهش از اجزاء متنوع آن، توصیفگر عبارتی (Term Compound) در برابر توصیفگر تک کلمه، تعداد واژه‌ها در توصیفگر عبارتی، توصیفگر در قالب صفت و موصوف در برابر توصیفگر با ترکیب حرف اضافه و یا حرف ربط، توصیفگر غیر مقلوب در برابر توصیفگر مقلوب (Inverted Descriptor)، استفاده از انواع علامت‌های نشانه‌گذاری در ساختار توصیفگر، استفاده از اختصار (Abbreviation) و آغاز

مواردی که در محدوده ۲۰ مقاله اول قرار گرفته بودند به عنوان مقالات مقیاس نهایی انتخاب شدند.

در کل ۱۶۷۹ مقاله مقیاس برای هر پایگاه اطلاعاتی جمع آوری شد که تعداد ۵۰۲ مقاله در زمینه بیماری های قلب و عروق، ۸۲۹ عنوان در زمینه سرطان ها، ۱۱۳ عنوان در زمینه بیماری های حول تولد و تعداد ۲۳۶ مقاله در زمینه بیماری های دستگاه تنفس بود. برای جمع آوری داده ها از لیست کنترل و پرسشنامه استفاده شد. تحلیل داده ها با استفاده از نرم افزار SPSS و آزمون های آماری t و Z انجام شد.

یافته ها

در تمامی زمینه های موضوعی تعداد توصیفگرهای ارائه شده در Emtree بیشتر از توصیفگرهای ارائه شده در MeSH می باشد. یافته ها نشان می دهد، در دو اصطلاحنامه توصیفگرها از نظر پراکندگی و تکرار در ساختار سلسله مراتبی، از الگویی متفاوت پیروی می کنند. میانگین تعداد تکرار درون موضوعی در MeSH (۷۵۴) و Emtree (۷۷۱) بوده است. آزمون اختلاف میانگین تکرار درون موضوعی در دو اصطلاحنامه نشان می دهد، با مقدار $t = -9/967$ و سطح معناداری $P = 0/000$ بین میانگین ها اختلاف معنی داری وجود دارد. میانگین تعداد تکرار برون موضوعی در MeSH (۷۷۰) و در Emtree (۲/۵۷) بوده است. آزمون اختلاف میانگین تکرار برون موضوعی در دو اصطلاحنامه نشان می دهد، با مقدار $t = -19/537$ و سطح معناداری $P = 0/000$ بین میانگین ها اختلاف معنی داری وجود دارد.

مقایسه بین دو اصطلاحنامه نشان می دهد که توصیفگرهای Emtree نسبت به MeSH بیشترین تعداد تکرار را در شاخه موضوعی اصلی خود و سایر شاخه های موضوعی داشته اند. میانگین تعداد مترادف ارائه شده در MeSH (۵ مترادف) و در Emtree (۲/۶ مترادف) می باشد. آزمون اختلاف میانگین تعداد مترادف های ارائه شده در دو اصطلاحنامه نشان می دهد، با مقدار $t = -10/092$ و سطح معناداری $P = 0/000$ بین میانگین ها اختلاف معنی داری وجود دارد و در نتیجه Emtree نسبت به MeSH تعداد مترادف بیشتری ارائه کرده است.

(Acronym) به عنوان توصیفگر، حروف تعریف در توصیفگر، توصیفگر با توضیحگر، ارائه یادداشت دامنه (Scope note) مورد ارزیابی قرار گرفته و سپس با استاندارد مقایسه شده است. علاوه بر این تعداد مترادف ها و تعداد تکرارهای بررسی شده است. برای بررسی شیوه های نمایه سازی نیاز به جمع آوری مقالات و طراحی راهبرد جستجوی مناسب بود که در هر دو پایگاه اطلاعاتی اجرا شود.

راهبرد جستجوی به کار برده شده عبارت بود از:

(A) AND [2006]/py AND [English]/limit
AND [humans]/limit AND [abstract]/limit=#1

سپس:

در پایگاه اطلاعاتی Pubmed: #1 AND "(B)"/[majr]/[mh]....
در پایگاه اطلاعاتی Embase: #1 AND "(B)"/exp/mj.....
(A) = ISSN مجله.

[2006]/py = تاریخ انتشار مقالات محدود به سال ۲۰۰۶ باشد.
[English]/limit = زبان مقاله محدود به انگلیسی باشد.
[humans]/limit = جامعه مورد بررسی در مقاله تنها انسانی باشد.

[abstract]/limit = مقالاتی که در پایگاه اطلاعاتی چکیده داشته باشند بازایی شود.
(B) = نام بیماری فرعی.

[majr] = توصیفگر بیماری فرعی در بین واژه های نمایه ای مقالات واژه نمایه ای اصلی باشد.
[mh] = توصیفگر بیماری فرعی در بین واژه های نمایه ای مقالات جزء واژه های نمایه ای فرعی هم باشد.

exp = به هنگام بازیابی اطلاعات علاوه بر توصیفگر بیماری فرعی، سایر توصیفگرهای اخص تر نسبت به آن را نیز بازیابی کند.

mj = توصیفگر بیماری فرعی در بین واژه های نمایه ای مقالات واژه نمایه ای اصلی باشد.

در هر پایگاه اطلاعاتی راهبرد طراحی شده ۴۸۱ مرتبه تکرار شد و پس از اجرای هر کدام از راهبردهای جستجو نتایج بازیابی شده بر طبق میزان ربط مرتب شدند. در دو پایگاه اطلاعاتی و در هر زمینه بیماری فرعی مقالاتی که برای هر عنوان مجله بازیابی شده بود با یکدیگر مقایسه گردید و

1. MeSH Major Topic
2. MeSH Terms
3. Explosion Search (Include sub-terms/derivatives)
4. Major Term

جدول ۱: نتایج حاصل از بررسی فاکتورهای مدنظر در دو اصطلاحنامه MeSH و Emtree و مقایسه با استاندارد ANSI/NISO Z39.19

ردیف	فاکتور بررسی شده در اصطلاحنامه	پیشنهاد استاندارد	نتایج آزمون‌های Z و t در دو اصطلاحنامه	وضعیت اصطلاحنامه MeSH	وضعیت اصطلاحنامه Emtree	اصطلاحنامه انطباق بیشتر با استاندارد
۱	توصیفگر عبارتی در برابر توصیفگر تک کلمه	ارجحیت توصیفگر عبارتی بر توصیفگر تک کلمه	اختلاف معنی دار وجود دارد	(۲۰/۴٪) توصیفگر تک کلمه	(۱۶/۴٪) توصیفگر تک کلمه	Emtree
۲	توصیفگر در قالب صفت و موصوف در برابر توصیفگر با ترکیب حرف اضافه و یا حرف ربط	ارجحیت توصیفگر در قالب صفت و موصوف بر توصیفگر با ترکیب حرف اضافه و یا حرف ربط	اختلاف معنی دار وجود دارد	(۲/۳٪) توصیفگر با ترکیب حرف اضافه یا حرف ربط	(۷/۳٪) توصیفگر با ترکیب حرف اضافه یا حرف ربط	Emtree
۳	تعداد واژه‌های توصیفگر عبارتی	ارجحیت در افزایش تعداد واژه‌های توصیفگر برای بیان مفاهیم بین موضوعی	اختلاف معنی دار وجود دارد	(۲/۹٪) میانگین تعداد واژه‌های توصیفگر عبارتی	(۲/۷٪) میانگین تعداد واژه‌های توصیفگر عبارتی	MeSH
۴	توصیفگر غیر مقلوب در برابر توصیفگر مقلوب	ارجحیت توصیفگر غیر مقلوب بر توصیفگر مقلوب	اختلاف معنی دار وجود دارد	(۲۳/۴٪) توصیفگر مقلوب	(۰/۰۳٪) توصیفگر مقلوب	Emtree
۵	استفاده از انواع علامت‌های نشانه گذاری	ارجحیت بر استفاده نکردن از انواع علامت‌های نشانه گذاری	اختلاف معنی دار وجود دارد	(۶۲/۵٪) توصیفگر بدون نشانه علامت گذاری	(۹۹/۷٪) توصیفگر بدون نشانه علامت گذاری	Emtree
۶	اختصار و آغاز و معادل آن به عنوان توصیفگر	استفاده از آنها با توجه به کثرت استفاده در متون رشته مجاز است	-	۲۶ توصیفگر اختصار یا آغاز و ۱۳۴ مترادف	۳۴ توصیفگر اختصار یا آغاز و ۱۳۸ مترادف	Emtree و MeSH
۷	استفاده از حروف تعریف در توصیفگر	ارجحیت بر استفاده نکردن از حروف تعریف در توصیفگر	اختلاف معنی دار وجود ندارد	۲ توصیفگر	۶ توصیفگر	-
۸	استفاده از توصیفگر با توضیحگر	استفاده از آن مجاز است ولی توصیفگر عبارتی بر توصیفگر با توضیحگر ارجحیت دارد	-	بدون توصیفگر	بدون توصیفگر	Emtree و MeSH
۹	ارائه یادداشت دامنه برای توصیفگر	ارجحیت بر ارائه یادداشت دامنه برای توصیفگر	اختلاف معنی دار وجود دارد	(۹۷/۲٪) یادداشت دامنه	(۵۷/۳٪) یادداشت دامنه	MeSH

X با توجه به ارجحیت توصیفگر عبارتی بر توضیحگر هر دو اصطلاحنامه با استاندارد مطابقت دارند.

در Pubmed، میانگین تعداد واژه‌های نمایه‌ای استخراج شده از عنوان (۶ واژه) است. در Embase میانگین تعداد واژه‌های نمایه‌ای استخراج شده از عنوان (۷ واژه) است. آزمون اختلاف میانگین بین تعداد واژه‌های نمایه‌ای استخراج شده از عنوان در دو پایگاه اطلاعاتی نشان می‌دهد با مقدار $(t = -10/114)$ و سطح معناداری $0/01$ ($P = 0/000$) بین میانگین‌ها اختلاف معنی داری وجود دارد و در نتیجه Embase نسبت به Pubmed مقالات را با تعداد واژه نمایه‌ای بیشتری از عنوان نمایه‌سازی می‌کند.

در Pubmed میانگین تعداد واژه‌های نمایه‌ای استخراج شده از چکیده ($10/4$ واژه) می‌باشد. در Embase میانگین تعداد واژه‌های نمایه‌ای استخراج شده از چکیده ($13/4$ واژه) می‌باشد. علاوه بر این آزمون اختلاف میانگین بین تعداد واژه‌های نمایه‌ای استخراج شده از چکیده در دو پایگاه اطلاعاتی نشان می‌دهد با مقدار $(t = -0/377)$ و سطح معناداری $0/05$ بین میانگین‌ها اختلاف معنی داری وجود ندارد و در نتیجه هر دو پایگاه اطلاعاتی از نظر تعداد واژه نمایه‌ای استخراج شده از چکیده مانند هم عمل می‌کنند.

نوع نمایه‌سازی Pubmed تخصیصی و در مقابل در Embase از نمایه‌سازی تخصیصی به همراه نمایه‌سازی استخراجی (زبان طبیعی) استفاده می‌شود.

برای ارزیابی کیفیت مقالات نمایه شده در دو پایگاه اطلاعاتی نیاز به نظرخواهی از پزشکان و تنظیم استاندارد طلایی مدنظر برای ارزیابی کیفیت نمایه‌سازی مقالات نمایه شده بود. در طراحی استاندارد طلایی چکیده و عنوان مقاله برای گروه شاهد و واژه‌های نمایه‌ای اصلی هر مقاله برای گروه آزمودنی تعریف گردید. سپس پرسشنامه‌ها در بین پزشکان دو بیمارستان حضرت رسول و بیمارستان امام خمینی توزیع شد. از ۱۶۷۹ پرسشنامه فرستاده شده ۱۱۱۶ ($66/4\%$) در صد پرسشنامه به طور کامل پر شده و برگشت داده شد. برای تمامی موارد مطرح شده در پرسشنامه‌ها پزشکان نظر موافق بر وجود ارتباط موضوعی بین واژه‌های نمایه‌ای اصلی با محتوای اطلاعاتی مقاله داشتند و تایید ارتباط موضوعی نشان دهنده کیفیت مناسب نمایه‌سازی مقالات است.

همان‌طور که جدول ۱ نشان می‌دهد، هیچ کدام از دو اصطلاحنامه انطباق ۱۰۰ با استاندارد نداشته‌اند. در ۴ فاکتور توصیفگر عبارتی در برابر توصیفگر تک کلمه، "توصیفگر در قالب صفت و موصوف در برابر توصیفگر با ترکیب حرف اضافه و یا حرف ربط"، "توصیفگر غیر مقلوب در برابر توصیفگر مقلوب" و "استفاده از انواع علامت‌های نشانه‌گذاری" Emtree تطابق بیشتری با استاندارد داشته است. در ۲ فاکتور تعداد واژه‌های توصیفگر عبارتی و ارائه‌یادداشت دامنه برای توصیفگر MeSH تطابق بیشتری با استاندارد داشته است. در فاکتورهای اختصار و آغاز و معادل آن به عنوان توصیفگر و استفاده از توصیفگر با توضیحگر، هر دو اصطلاحنامه مطابق با استاندارد بوده‌اند و در آخر در فاکتور استفاده از حروف تعریف در توصیفگر هیچ‌کدام از دو اصطلاحنامه مطابق با استاندارد نبوده‌اند. قابل ذکر است که در نمونه جمع‌آوری شده از دو اصطلاحنامه، توصیفگر با توضیحگری وجود نداشته است. مقایسه بین دو اصطلاحنامه نشان می‌دهد، Emtree در ۶ مورد و MeSH نیز در ۴ مورد انطباق بیشتری با استاندارد داشته‌اند.

در بررسی شیوه‌های نمایه‌سازی در دو پایگاه اطلاعاتی برای شمارش تعداد واژه‌های نمایه‌ای (Indexing Terms) داده شده به هر مقاله تمامی انواع واژه‌های نمایه‌ای شمارش گردید. بدین ترتیب که در Pubmed تمامی واژه‌های نمایه‌ای داده شده برای نوع مقاله (Publication Types) واژه‌های MeSH و نوع مواد (Substances) به کار رفته در تحقیق شمارش شد و در Embase نیز واژه‌های نمایه‌ای دارویی شامل دو گروه اصلی و فرعی و واژه‌های نمایه‌ای بالینی شامل دو گروه اصلی و فرعی شمارش گردید.

در Pubmed میانگین تعداد واژه‌های نمایه‌ای ارائه شده به مقالات ۲۵ واژه نمایه‌ای و در Embase ۳۳ واژه نمایه‌ای می‌باشد. آزمون اختلاف میانگین تعداد واژه نمایه‌ای ارائه شده به مقالات در دو پایگاه اطلاعاتی نشان می‌دهد با مقدار $(t = -42/987)$ و سطح معناداری $0/01$ ($P = 0/000$) بین میانگین‌ها اختلاف معنی داری وجود دارد و در نتیجه در Embase نسبت به Pubmed تعداد واژه نمایه‌ای بیشتری به مقالات اختصاص داده شده است.

بحث و نتیجه‌گیری

در Emtree به نسبت MeSH از تعداد توصیفگر بدون مترادف کمتری استفاده شده است. هر چه تعداد مترادف‌های ارائه شده برای توصیفگر بیشتر باشد، کاربران به هنگام بازیابی اطلاعات نیاز کمتری به دانستن واژه انتخاب شده توصیفگر خواهند داشت. در ۴ زمینه موضوعی مدنظر پژوهش، تمامی توصیفگرهای Emtree توصیفگرهای معادل از MeSH را در بین مترادف‌های خود دارند. بنابراین کاربران به راحتی می‌توانند با استفاده از توصیفگرهای MeSH در پایگاه اطلاعاتی Embase نیز جستجو کنند و به هنگام جستجو از Pubmed به Embase دچار مشکل نمی‌شوند.

توصیفگرها در Emtree به نسبت MeSH در تعداد دفعات بیشتری تکرار شده‌اند. تکرار بیشتر یک توصیفگر در ساختار اصطلاحنامه موجب می‌شود که توصیفگر رابطه محتوایی گسترده‌تری با موضوعات مختلف در اصطلاحنامه داشته باشد. از ۹ مورد بررسی شده، ۸ مورد مربوط به ساختار واژگانی توصیفگرها است و یک مورد باقی‌مانده (یادداشت دامنه) درباره اطلاعات تکمیلی است که برای یک توصیفگر فراهم می‌شود. در ۸ مورد بررسی شده، MeSH در ۳ فاکتور و Emtree در ۶ فاکتور تطابق بیشتری با استاندارد داشته‌اند.

یافته‌های حاصل از بررسی فرضیه پژوهش و مطلب فوق نشان می‌دهد، برای الگوبرداری اصطلاحنامه پزشکی فارسی، ساختار سلسله‌مراتبی و ترکیب بین موضوعی واژگان توصیفگرهای اصطلاحنامه Emtree مناسب‌تر است.

با توجه به میانگین واژه‌هایی که در هر پایگاه اطلاعاتی به مقالات داده می‌شود، می‌توان نتیجه گرفت که نمایه‌سازی Embase جامع‌تر از نمایه‌سازی Pubmed می‌باشد. علاوه بر این تعداد واژه‌های نمایه‌ای بیشتر نشان‌دهنده جزء نگرتر بودن [۵] نمایه‌سازی Embase نسبت به Pubmed است. پایگاه اطلاعاتی که از نمایه‌سازی جامع استفاده می‌کند، به هنگام جستجوی اطلاعات تعداد مدارک بیشتری را بازیابی می‌کند. هر چه نمایه‌سازی گزیده‌تر باشد، برای طراحی راهبرد جستجو استفاده از عملگر "یا" (OR operator) تعداد مقالات بازیابی شده را افزایش می‌دهد. در مقابل هر چه در نمایه‌سازی

تعداد واژه‌های نمایه‌ای بیشتری به مقالات اختصاص داده شود، استفاده از عملگر "و" (AND operator) مقالات مرتبط‌تر با موضوع مدنظر را بازیابی می‌کند. [۵] راهبرد جستجویی که در ترکیب واژه‌های آن از عملگر "یا" استفاده شده باشد مناسب برای پایگاه اطلاعاتی Pubmed است و راهبرد جستجویی که در ترکیب واژه‌های آن از عملگر "و" استفاده شده باشد مناسب برای پایگاه اطلاعاتی Embase است.

در Pubmed، به طور متوسط ۲۵ واژه نمایه‌ای به هر مقاله داده می‌شود که از این تعداد، ۶ واژه از عنوان و ۱۰/۴ واژه از چکیده مقاله استخراج می‌شوند. ۸ واژه نمایه‌ای باقی‌مانده از سایر بخش‌های مقاله استخراج می‌گردند.

در Embase، به طور متوسط ۳۳ واژه نمایه‌ای به هر مقاله داده می‌شود که از این تعداد، ۷ واژه از عنوان و ۱۳/۴ واژه از چکیده مقاله استخراج می‌شوند. ۱۲ واژه نمایه‌ای باقی‌مانده از سایر بخش‌های مقاله استخراج می‌گردند.

بنابراین می‌توان نتیجه گرفت، هر دو پایگاه اطلاعاتی تمامی متن مقالات را در نمایه‌سازی در نظر دارند. با توجه به نتایج فوق و اینکه در نمایه‌سازی پایگاه‌های اطلاعاتی ایرانی واژه‌های نمایه‌ای بیشتر از عنوان مقاله استخراج می‌شوند لذا پیشنهاد می‌شود که نمایه‌سازان به هنگام نمایه‌سازی سایر قسمت‌های مقاله را نیز در نظر داشته باشند. ♦

فهرست منابع

۱. نقوی محسن. سیمای مرگ و میر در ۲۳ استان کشور در سال ۱۳۸۲. معاونت سلامت وزارت بهداشت درمان و آموزش پزشکی، تهران، ۱۳۸۴.
2. PubMed Overview. Available at: [overview.html/http://www.ncbi.nlm.nih.gov/entrez/query/static/overview.html](http://www.ncbi.nlm.nih.gov/entrez/query/static/overview.html)
3. About Embase. Available at: <http://www.info.embase.com/embasesuite/about/>
۴. اسدی اربابی عاطفه. بررسی رفتارهای اطلاع‌یابی محققان مراکز تحقیقاتی علوم پزشکی شهر تهران. پایان‌نامه کارشناسی ارشد رشته کتابداری و اطلاع‌رسانی پزشکی، تهران: دانشکده مدیریت و اطلاع‌رسانی پزشکی دانشگاه علوم پزشکی ایران، ۱۳۸۴.
۵. لنکستر فردریک ویلفرید. نمایه‌سازی و چکیده‌نویسی، مبانی نظری و عملی. ترجمه گیلوری عباس. چاپار، تهران، ۱۳۸۲.

Structures of Pubmed and Embase databases with NISO standard of the thesauri; to assessment of these databases' indexing methods

Khani S.¹ / Alibeig M. R.² / Haghani H.³

Abstract

Introduction: According to mortality rates in Iran, cardiovascular diseases, neoplasms, perinatal mortality, and respiratory tract diseases were top rate mortality in 2003(1382). To reduce mortality rate, Iranian medical community need to know more about recent therapeutic regimens. Two main medical databases are Pubmed and Embase. Researching Pubmed and Embase indexing methods and comparing MeSH with Emtree help users to do more successful search in these databases. Consequently, designers of national medical information database in Iran may construct a model for updating indexing methods and thesaurus. This study aimed at comparing indexing methods in Pubmed and Embase.

Methods: This was an applied; descriptive - analytical research. Research population was all of the descriptors in MeSH and Emtree; and indexed articles from Pubmed and Embase about four selected fields. In the last 3 months in 2006, all of the descriptors of selected fields were extracted through a structured search strategy. Then needed data was extracted from 6381 descriptors and 3358 articles. For collecting data we used a checklist and a questionnaire. Nine factors (including phrased descriptors versus single word descriptors, number of words in phrased descriptor, descriptor on adjectives and substantives format versus prepositions and conjunctives format, transforming versus non-transforming descriptors, using different quotation sign in descriptor structure, using abbreviations and commencer as descriptor, using definitions in descriptors, descriptors with explanations, and providing comments) selected from standard and analyzed in thesauri. Data were analyzed by SPSS using t-test and z test.

Results: Emtree in six factors, and MeSH in four factors are more similar to standards. Pubmed articles are indexed with average number of 21-30 indexing terms. Embase uses average number 31-40 indexing terms for each article.

Conclusions: Emtree structure is more suitable for modeling. Embase indexing method is assignmotive and derivative indexing and does it specific and more exhaustive. Pubmed indexing method is derivative and exhaustive indexing.

Keywords: *MeSH, Emtree, Thesauri assessment, Pubmed, Embase, Indexing methods.*

1. M.S in Medical Information and Library Science, School of Management and Medical Information Sciences ,Iran University of Medical Sciences

2. Faculty member, School of Management and Medical Information Sciences, Iran University of Medical Sciences

3. Faculty member, School of Management and Medical Information Sciences, Iran University of Medical Sciences